

学校编码: 10384

学号: 15420121151879

分类号 _____ 密级 _____

UDC _____

廈門大學

硕 士 学 位 论 文

数据挖掘在资金中介筛选中的应用研究

Application Research of Data Mining in Financial Intermediary Screening

梁俊憬

指导教师姓名: 郭 鹏 辉 副教授

专 业 名 称: 数 量 经 济 学

论文提交日期: 2015 年 3 月 日

论文答辩时间: 2015 年 5 月 日

学位授予日期: 2015 年 月 日

答辩委员会主席: _____

评 阅 人: _____

2015 年 月 日

厦门大学学位论文原创性声明

本人呈交的学位论文是本人在导师指导下，独立完成的研究成果。本人在论文写作中参考其他个人或集体已经发表的研究成果，均在文中以适当方式明确标明，并符合法律规范和《厦门大学研究生学术活动规范（试行）》。

另外，该学位论文为()课题(组)的研究成果，获得()课题(组)经费或实验室的资助，在()实验室完成。（请在以上括号内填写课题或课题组负责人或实验室名称，未有此项声明内容的，可以不作特别声明。）

声明人（签名）：

2015 年 月 日

厦门大学学位论文著作权使用声明

本人同意厦门大学根据《中华人民共和国学位条例暂行实施办法》等规定保留和使用此学位论文，并向主管部门或其指定机构送交学位论文（包括纸质版和电子版），允许学位论文进入厦门大学图书馆及其数据库被查阅、借阅。本人同意厦门大学将学位论文加入全国博士、硕士学位论文共建单位数据库进行检索，将学位论文的标题和摘要汇编出版，采用影印、缩印或者其它方式合理复制学位论文。

本学位论文属于：

（ ） 1.经厦门大学保密委员会审查核定的保密学位论文，
于 年 月 日解密，解密后适用上述授权。

（ ☒ ） 2.不保密，适用上述授权。

（请在以上相应括号内打“√”或填上相应内容。保密学位论文应是已经厦门大学保密委员会审定过的学位论文，未经厦门大学保密委员会审定的学位论文均为公开学位论文。此声明栏不填写的，默认为公开学位论文，均适用上述授权。）

声明人（签名）：

2015 年 月 日

摘要

随着我国经济不断的发展，居民的收入不断的提升，居民手中的闲余资金也随之增长。但是由于我国资本市场的不完善，所能提供的金融工具品种单一，导致居民的投资渠道相对匮乏，这就促使了资金中介的产生与发展。所谓资金中介是指不受国家有关部门以及相关法律法规监管的，为民间存贷款以及民间投资融资提供相应渠道的个人或机构。随着资金中介规模的不断壮大，其自身存在的问题逐渐暴露出来。目前，国内外学者只是从资金中介的定义、资金中介的形成原因以及国内资金中介所存在的问题这三个方面对资金中介进行研究分析，而并没有描述资金中介的具体特征。并且现今许多商业银行也只是通过人工甄别的方法筛选资金中介，即通过业务人员的工作经验去识别资金中介，这种方法效率低下，且准确率也不高。因而本文希望通过利用数据挖掘技术，建立能够反映资金中介主要行为特征的模型，以期通过该模型提高资金中的筛选效率与准确率，对资金中介实施有效的监管。

本文先经过理论分析、实证模拟以及实务论证这三个方面的流程，确定本文所要建立的模型，同时有效地扩充了业务提供的资金中介客户名单，并基于扩充名单进行数据准备和数据分析，建立决策树模型、变量聚类模型以及逻辑斯蒂回归模型，最后将逻辑斯蒂回归模型的结果应用于业务验证。基于上述筛选过程，本文研究发现资金中介呈现如下几个特征：（1）资金中介的年龄大多分布在 30 岁至 40 岁之间；（2）资金中介多有金融行业的从业背景；（3）资金中介之间相互分工和相互合作的现象比较普遍；（4）资金中介的交易量与交易金额呈现出逐年递增的趋势；（5）亲戚共同成为资金中介的现象十分常见。基于资金中介的发展背景以及上述 5 个结论，本文提出正确引导资金中介的几点建议：（1）进一步完善我国的收入分配结构；（2）加强金融行业对于自身客户信息的监管力度；（3）完善有关资金中介的法律体系；（4）进一步完善我国资本市场；（5）加大宣传资金中介在我国经济体系中的重要地位。

关键词：数据挖掘；资金中介；客户筛选

Abstract

With development of economy and constant promotion of income, spare funds of residents increases. However, due to imperfect capital market, it can only provide a few financial instruments, which leads to lack of investment channels for residents, and prompts the emergence and development of financial intermediary. The so-called financial intermediary means the individual or institution who provides the appropriate channels for civil deposits and loans as well as private investment financing without supervision of the relevant laws, regulations and national authorities. As the scale of financial intermediary increasing, its own problems gradually expose. Currently, many scholars do research into financial intermediary only from its definition, the causes of its emergency and its problems, and they do not describe specific characteristics of financial intermediary. And many commercial banks sift financial intermediary only by artificial screening, that is, screening financial intermediary by the work experience of staff. This approach is inefficient and the accuracy of it is not high. Therefore this paper hopes to establish a model which can reflect the main characteristic of financial intermediary by using technique of data mining, in order to enhance efficiency and accuracy of financial intermediary screening by applying the model, and implement effective supervision for financial intermediary.

First, this paper determines the model which will be used later according to theoretical analysis, empirical stimulation and practice verification, then expands the sample supplied by business, conducts data preparation, data analysis, and establishes decision-tree model, variable clustering model and logistic regression model based on the sample, finally, carries out business validation for the results of logistic regression model. Based on the screening process, this paper finds several characteristics of financial intermediary, that is, 1. The age of most of financial intermediary is between thirty and forty; 2. Most of financial intermediary have finance background; 3. The phenomenon of cooperation among financial intermediary is very general; 4. Trading volume and trading amount of financial intermediary emerges the trend of increasing

with years; 5. The phenomenon that relatives are all financial intermediary is very general. Based on development background of financial intermediary and the five conclusions described above, this paper puts forward some advice: 1. Further improve the income distribution structure; 2. Strengthen the supervision of customer information of the financial sector; 3. Improve the legal system concerning financial intermediary; 4. Further improve the capital market; 5. Publicize the important role of financial intermediary in economic system.

Key Words: Data Mining; Financial Intermediary; Customer Screening

目录

第 1 章 绪论	1
1.1 资金中介发展背景	1
1.2 文献综述	2
1.2.1 资金中介的相关研究	2
1.2.2 数据挖掘技术的应用现状	7
1.3 本文研究贡献	10
1.4 本文框架	10
第 2 章 模型介绍	11
2.1 聚类模型	11
2.2 决策树模型	12
2.3 逻辑斯蒂回归模型	14
2.4 本章小结	14
第 3 章 数据挖掘在资金中介筛选中的应用研究	16
3.1 资金中介项目介绍	16
3.2 资金中介业务现状分析	16
3.3 数据挖掘应用研究与分析	17
3.3.1 模型确定	17
3.3.2 样本分析与扩充	21
3.3.3 数据准备	22
3.3.4 数据分析	24

3.3.5 模型建立与比较	27
3.3.6 资金中介筛选流程小结	36
3.3.7 业务验证	37
3.4 本章小结.....	38
第 4 章 研究总结.....	40
4.1 研究总结.....	40
4.2 正确引导资金中介的几点建议.....	41
4.3 本文研究不足.....	42
参考文献.....	43
致谢语.....	47

Contents

Chapter1 Introduction.....	1
1.1 Development Background of Financial Intermediary	1
1.2 Literature Review	2
1.2.1 Research of Financial Intermediary	2
1.2.2 Status of Application of Data Mining	7
1.3 Contribution.....	10
1.4 Frame.....	10
Chapter2 Introduction of Model.....	11
2.1 Clustering Model.....	11
2.2 Decision-tree Model.....	12
2.3 Logistic Regression Model.....	14
2.4 Summary of this Chapter.....	15
Chapter3 Application Research of Data Mining.....	16
3.1 Introduction of Financial Intermediary.....	16
3.2 Status Analysis of Financial Intermediary.....	17
3.3 Application Research and Analysis of Data Mining.....	17
3.3.1 Model Determining	20
3.3.2 Analysis and Expansion of Sample	22
3.3.3 Data Preparation	23
3.3.4 Data Analysis	27

3.3.5 Establishment and Comparison of Model.....	36
3.3.6 Summary of Financial intermediary Item.....	37
3.3.7 Business Validation	38
3.4 Summary of this Chapter.....	40
Chapter4 Conclusion	40
4.1 Conclusion.....	43
4.2 Advice.....	41
4.3 Deficiency.....	42
References	43
Acknowledgements	47

第 1 章 绪论

本文将利用数据挖掘技术进行资金中介的筛选，在此之前，本文先介绍资金中介的发展背景及其相关研究分析。

1.1 资金中介发展背景

随着我国综合国力的不断提高，我国居民的收入也有了显著的提升，城镇居民人均可支配收入在 2004 年的时候为 17901.06 元，而到了 2011 年便上升到了 41438.58 元。收入提升了，居民手中的闲余资金也增多了，但是由于我国金融市场的不完善，市场上所提供的金融工具品种单一，导致居民的投资渠道相对匮乏。而近些年来我国的股票市场与房地产市场动荡不定，投入这两个市场的资金也逐渐变少，出于资金高回报率的诱惑，许多居民将手中的闲钱投向了民间金融市场、而同时我国银行的贷款审批周期较长，再加上我国银行存在“信贷歧视”的现象，即偏向于放贷给规模较大的企业、跨国集团、国有企业以及国家积极扶持的相关行业，这样便导致急需资金的中小型企业与广大居民不能及时得到所需的资金，无可奈何之下，只能转向民间金融市场，资金中介就是在这样的环境下逐渐发展壮大起来。所谓资金中介就是不受政府相关部门监管的，为民间的借款贷款、投资融资提供中介服务的个人或机构，目前资金中介存在的组织形式主要有亲友或企业间偶然借贷、联结贷款、专业的存款收集者和放贷者、合会、典当行、私募资本市场、地下银行或地下钱庄、小额信贷组织、合作金融组织以及民间票据市场等。中央财经大学课题组（2004）指出，中国地下信贷规模介于 7400 亿至 8400 亿之间，而根据 Yuande Zhu, Helen Qiao and Ernest Ho（2011）的调查发现，中国民间信贷规模约为 3.8 万亿，由此可见，资金中介的规模在不断地壮大，并且有继续壮大的趋势。但是，近些年来出现的“温州跑路潮”、集资诈骗以及集资出逃等事件将资金中介本身存在的问题、风险逐步暴露出来，给人们敲响了警钟，促使人们加强对资金中介的监管力度。

1.2 文献综述

第二节中将介绍关于资金中介的相关研究分析、数据挖掘技术的应用现状以及指出数据挖掘技术在资金中介筛选中的适用性。

1.2.1 资金中介的相关研究

国内外学者对于资金中介的定义、资金中介的形成原因以及国内资金中介发展中所存在的问题进行了相关的研究分析，下面本文将从资金中介的定义、资金中介的形成原因以及国内资金中介发展中所存在的问题这三个方面来对国内外学者的研究分析进行归纳总结。

1. 资金中介的定义

目前在学术界，对于资金中介的定义并没有统一的标准。Gurley J. G. and Shaw E. S. (1956) 将资金中介定义为不受中央银行和国家法律法规监督管理的金融中介，其一般不受中央银行所制定的法定准备金制度、资产负债比制度以及资本充足率制度的限制，而且资金中介的经营活动也不需要向中央银行进行汇报，其可以自主地决定利率，并且资金中介所确定的利率往往要高于正规金融机构所制定的利率。Fry R. G. (1982) 提出资金中介是指不受政府有关部门以及相关法律法规对于资本金、存贷款利率限制、储备金和资金流动性、强制性信贷对象以及定期提交审计报告等要求束缚、为个人与个人或者个人与私营企业主提供间接投资或者直接投资的中介人或中介机构，通常并没有得到政府有关部门以及相关法律法规的认可，一般又被称为非制度金融中介、民间金融中介、地下金融中介以及草根金融中介等等，资金中介经常处于非法的阴影笼罩之下。Gurley and Shaw (1984) 则认为资金中介是游离于正规金融系统之外的，专门从事不受国家以及中央银行管控的存款、贷款以及其他形式的资金交易活动的个人或者机构。谈儒勇 (2001) 指出，资金中介是指非正式的、不受法律法规管制的、尚未在工商局登记的、未被记录的非正规金融。王爱俭, 张全旺以及于学伟 (2004) 将资金中介定义为在发展中国家，专门为非法人组织（如小农场主、中小企业以及零售商等）提供贷款的专业放款人、商人或者典当商。张杰 (2006) 则提出资金中

介是指专门以相关法律法规未规定的形式从事资金融通业务的中介服务人,通常包括地下金融以及民间金融。

综上所述,本文将资金中介定义为不受国家有关部门以及相关法律法规监管的,为民间存贷款以及民间投资融资提供相应渠道的个人或机构,该定义与大多数学者对于资金中介的定义均相近。

2. 资金中介形成的原因

目前国内外学者关于资金中介形成的原因主要有三个不同的观点:(1)金融抑制假说,该假说指出由于正规金融机构实行信贷配给(根据 Smith (1976) 对于信贷配给的定义,从理论上来说,信贷配给是指在一定的利率之下,由于市场上对于贷款的需求远远超出了资金供给,导致市场不能出清,信贷配给的具体表现为资金短缺者希望从正规金融机构获取所需的资金,而正规金融机构由于放贷资金不足而无法提供该资金短缺者所需的资金),正规金融机构只将资金放贷给国家希望优先发展的行业,而其余的行业得不到充分的资金支持,只能从民间金融市场寻求所需的资金,资金中介由此发展起来;(2)交易成本假说,该假说指出由于正规金融机构对于大多数中小企业以及广大民众的财务状况、个人道德水平以及偿还债务的能力均没有充分的了解,因而正规金融机构出于合同执行成本和违约风险的考虑,不太愿意放贷给这部分群体,而大型企业将自己的各种信息对外充分披露,正规金融机构可以根据这些信息判断是否可以将资金放贷给这些企业,再加上放贷给这些企业的贷款具有规模效应,如此一来,大大降低了正规金融机构所需付出的交易成本,因而正规金融机构比较倾向于放贷给大型企业,而急需资金的中小企业与广大民众只能依靠资金中介获取他们所需的资金;(3)信息保密假说,该假说指出由于正规金融机构的交易行为是建立在信息充分披露的基础之上,而有些客户出于自身信息的安全,因而这部分客户不愿意通过正规金融机构进行金融交易,比如部分金融交易中可能存在违法行为,这部分交易若是通过正规金融机构来完成,其所面临的风险较大,其中的违法行为极有可能被发现,进而受到法律的制裁,从而这部分交易通常是通过资金中介来完成,资金中介就是这样形成的。下面本文将从金融抑制假说、交易成本假说以及信息保密假说这三个原因来归纳国内外学者对于资金中介形成原因的研究分析。

（1）金融抑制假说

McKinnon R. I. (1973) 指出正是由于国家过度得干预资本市场才导致金融抑制的形成，同时过低的市场利率，使得市场上的资金需求超出了资金供给，正规金融机构才不得不采取信贷配给，只将资金放贷给拥有特权的客户以及国家优先扶植的企业，比如大型跨国公司、国家控股的公司等等，而其余的客户只能从资金中介手中得到他们所需的资金。

易秋霖与郭慧（2003）指出，除了那些从事非法经济活动的资金中介，比如利用地下钱庄进行洗钱等，其余资金中介形成的原因主要是金融抑制，而金融抑制主要包括对正规金融机构信贷额度的计划控制、对设立非公有制金融机构的限制、对正规金融机构存款利率的计划控制、对正规金融机构经营领域的限制以及对正规金融机构向中小企业主贷款的限制等等。

张杰（2006）利用借贷资金理论模型研究可贷资本市场上的资金供求关系，并且指出，政府为了刺激广大民众的投资，拉动经济增长，采取了低利率政策，将市场利率定在一个较低的水平之上，这使得可贷资本市场上的资金需求大幅上升，资金供给反而大幅下降了，形成了巨大的资金缺口，从而导致了金融抑制，大大地限制了民众的投资，而资金中介就是在这样的环境下产生并且逐步发展起来。

（2）交易成本假说

交易成本假说是基于信息不对称提出的，金融交易中的交易成本主要是由以下三个方面构成的：（1）在签订贷款合约之前，金融机构需要对贷款者的财务状况、信用度以及偿还债务的能力进行详细地调查，其中所需的调查成本构成了交易成本的一部分；（2）签订贷款合约之后对贷款者的使用情况的监督成本；（3）贷款合约到期之后，为了保证贷款者按期偿还所贷的款项以及利息，或者贷款者不能按期偿还本金与利息之后通过法律途径或其他相关方式使得贷款者履行贷款合约的合同执行成本。

Gurley J. G. and Shaw E. S. (1956) 认为，由于资金中介的放贷对象通常都是他们所熟悉的人或者是亲戚，因而对放贷对象的债务偿还能力比较了解，这样就顺利解决了信息不对称问题，相比于正规金融机构，这是资金中介最大的优势。

Degree papers are in the “[Xiamen University Electronic Theses and Dissertations Database](#)”. Full texts are available in the following ways:

1. If your library is a CALIS member libraries, please log on <http://etd.calis.edu.cn/> and submit requests online, or consult the interlibrary loan department in your library.
2. For users of non-CALIS member libraries, please mail to etd@xmu.edu.cn for delivery details.

厦门大学博硕士论文摘要库